# Chemical protein synthesis: from nanobodies to activity-based probes

Huppelschoten, Y.

# CHAPTER 8

## Summary and Future Perspectives

Access to site-specifically modified proteins unattainable via biochemical or enzymatic approaches has greatly increased over the past few decades due to the improvement in chemical protein synthesis. These synthetic proteins have contributed to the fundamental understanding of various biochemical processes and structure elucidation.[1] The success of chemical protein synthesis is based on the revolutionary progress made in the peptide chemistry field. Solid phase peptide synthesis (SPPS) has enabled the linear synthesis of proteins over 100 amino acid residues.[2] In addition, the development of native chemical ligation (NCL); the chemo selective reaction between two unprotected peptides to form a native peptide bond has increased the accessibility of synthetic proteins.[3] The stepwise improvement and discovery of ligation techniques and SPPS over the past years are discussed in **chapter 1** of this thesis.

With this progress in chemical protein synthesis, the amount of chemically synthesized protein derivatives used to elucidate biological processes greatly expanded. One research field that benefited from this is ubiquitin (Ub) and ubiquitin-like (Ubl) signaling. Chemical tools modified with valuable chemical moieties e.g. fluorescent tags or warheads have contributed to the deciphering of Ub signaling. Especially, the activity-based probes (ABPs) containing a propargyl (PA) moiety, capable of selectively capturing cysteine proteases have contributed greatly to the field.[4–6] In **chapter 2**, the progress in the chemical protein synthesis of state-of-the-art Ub and Ub-like chemical probes, their unique concepts and related discoveries in the Ub field are discussed. Next, in the experimental chapters of this thesis, we set out to apply SPPS and chemical protein synthesis in the preparation of novel assay-reagents and ABPs.

Nanobodies (Nbs) are single antigen binding variable heavy-chain domain (VHH) produced by camelid species, which are seen as promising therapeutics.[7,8] Accordingly, interest has been raised for the functionalization of Nbs for various applications such as diagnostic tools, Nb-drug conjugates, and bivalent Nb conjugates.[9–11] We imagined that, the chemical synthesis of a Nb could accelerate the process of generating homogeneous Nb-conjugates. Diverse functional groups suitable for the chemo-selective labeling can be easily introduced at defined, non-interfering regions of the Nb through a synthetic approach. The chemical synthesis of a functionalized Nb targeting green fluorescent protein (GFP) using a three-segment NCL approach is presented in **chapter 3**. A C-terminal propargyl moiety for selective on demand functionalization using copper mediated azide alkyne cycloaddition chemistry (CuAAC) was introduced. The propargyl moiety is installed into the Nb sequence without compromising the antigen-binding sites, hence leaving the Nbs binding affinity unaffected. Functionalization of the Nb with an affinity tag or fluorescent dye resulted in two distinct Nb probes which were successfully validated in *in vitro* assays (Fig. 1). Furthermore, comparison of the synthetic GFP-Nb to an expressed GFP-Nb in biophysical experiments provided similar binding

affinities and folding for both functionalized and non-functionalized synthetic Nbs. In general, Nbs share a highly conserved structure, therefore the presented methodology could pave the way for a modular synthetic approach towards this class of proteins. The advantage of using SPPS is the easy introduction of unnatural amino acids, including amino acids stabilizing the protein against *in vivo* degradation.[12] Future endeavors using our chemical synthesis and on-demand functionalization approach might yield even more sophisticated Nb-conjugates, carrying a wide variety of payloads that could be targeted to specific locations in tissues. Hence, expansion of this methodology has potential in e.g. the streamlined preparation of promising and more stable therapeutic Nb-drug conjugates or Nb multimers.

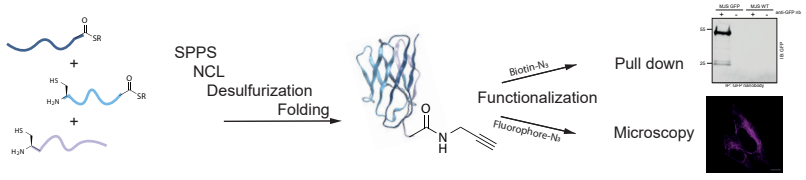


**Figure 1.** Schematic representation of the synthesis and validation of a synthetic GFP Nb.

Small ubiquitin-modifier (SUMO) is a post-translational modification, that in analogy to Ub is attached to (mostly) lysine residues in target proteins. SUMOylation affects over a thousand proteins in mammalian cells and is thereby a key event in various biological processes e.g. protein stability, cell stress responses and transcription factor activity.[13] Hence, SUMOylating ligases as well as deSUMOylating enzymes (sentrin-specific proteases (SENPs)) play a critical role in maintaining cellular physiology. The SENP cysteine proteases are less studied when compared to deubiquitinating proteases (DUBs). To shed light on the still enigmatic cleavage specificity and substrate selectivity of SENPs, a broad set of SUMO based iso-peptide linked assay reagents was envisioned. Isopeptide-linked Ub(l)-based fluorescent polarization reagents have been applied in several studies including the study of the proteolytic preference of SARS-CoV2 protease[14] and to elucidate enzyme mechanisms,[15] thereby successfully contributing in the elucidation of Ub signalling and specificity of deubiquitinating enzymes (DUBs). The synthesis of SUMO fluorescent polarization (FP) probes in 96-well format, based on a native chemical ligation – desulfurization approach and the subsequent use in a FP assay towards several SENPs is described in **chapter 4** (Fig. 2). The parallel synthesis protocol for NCL between γ-thiolysine containing peptides and SUMO3-thioester in a 96-well plate was optimized so an one-pot ligation and desulfurization could be performed. The native cysteine residue of SUMO3 could be preserved by performing the NCL and desulfurization under native conditions and hence the need for orthogonal protection of this cysteine and an elaborate deprotection scheme was avoided. After synthesis and HPLC-purification 83 SUMO based assay reagents were obtained, showcasing the robustness of this approach. Next, five SENPs were screened for their (specific) deconjugating activity of the FP-probes. Different intrinsic activities were found between the SENPs and preferences

for specific subsets of probes were observed. Some of this proteolysis-data is in line with literature reports, whereas other findings pose interesting questions to yet unallocated targets for specific SENPs. Of note is that SENP activity was measured on SUMO-peptide conjugates that might not completely correlate to the corresponding SUMO-protein substrates found *in vivo*. Hence, we tried to validate some of the results obtained from FP-reagents derived from diSUMO linkages on chemically prepared native full length di-SUMO proteins. Six native diSUMOs were prepared using the synthesis protocol for the probes preventing the native cysteine from desulfurization. Initial gel based cleavage assays of the diSUMOs displayed comparable results as in the FP-assay, however further validation is needed. The addition of a fluorescent tag to the diSUMO probes could increase visualization sensitivity and thereby clarify SENPs preference. It should also be mentioned that several FP-probes are derived from the same protein target, however from different lysine residues possibly casting a light on the preferred sites of SUMO deconjugation by the assayed SENP. For example, there are three probes derived from the target protein promyelocytic leukemia protein (PML); lysine490, lysine160, and lysine380 . Interestingly, SENP6 efficiently processes the probe derived from lysine160 while lysine490 and lysine380 are processed significantly less. It is known that SENP6 is responsible for the deSUMOylation of PML, however which residues are unknown.[16] Site specific auxiliary-meditated SUMOylation of PML on the mentioned lysine residues could clarify if SENP6 really has a preference for lysine180. Like PML, there are several other proteins (PARP1, PIAS1 and TRIM28 )with different probes and interesting SENP preference profiles. Finally, access to SENP3 would complete the panel of SENPs and could potentially add valuable data to this already extensive data set, however recombinant expression of this SENP is still problematic. All in all **chapter 4** shows that in-plate synthesis of SUMO-FP reagents yields a functional assay that serves as an excellent starting point for SENP profiling, although careful analysis and future validation using orthogonal assays are needed to fully understand deSUMOylating enzymes in detail.

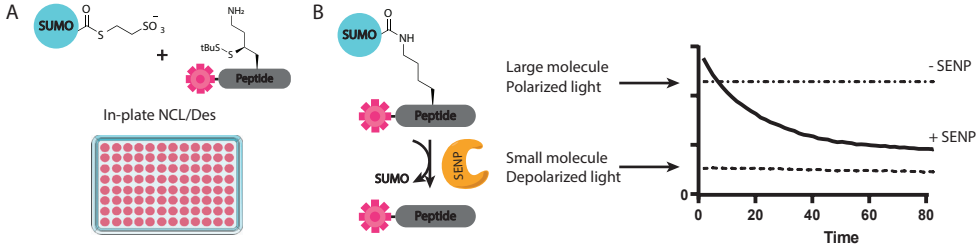


**Figure 2. A.** Schematic representation of native chemical ligation and desulfurization in 96-well format. **B.** FP assay. When a fluorophore, covalently attached to a small peptide is excited by polarized light, it will mainly emit depolarized light. When it is attached to a larger protein (e.g. SUMO) the emitted light is less depolarized. The activity of the enzyme can be monitored by following the change in fluorescence polarization.

Another understudied Ubl family of interest are the autophagy-related (Atg) proteins, or more specifically microtubule-associated protein light chain 3 A (LC3A) and B (LC3B).[17] LC3B is a key component in the degradation of intercellular materials during a process called autophagy. LC3B is conjugated to phospholipids and incorporated in double-membrane vesicles (autophagosomes) responsible for the engulfment of components ready for degradation.[18] Dysregulation of this process has been associated with a variety of human diseases, hence components of the autophagy pathway are considered as potential therapeutic targets since.[19] The mechanistic details surrounding the biological function of LC3 specific enzymes, however is far from unraveled, partly due to the lack of (chemical) tools for activity profiling. To gain access to LC3 based activity based probes (ABPs) several synthetic routes are explored in this thesis. In chapter 5 a step-by-step protocol for the linear synthesis of LC3A and LC3B containing a fluorescent label using infrared heating is described. Both proteins are 120 amino acids long and were successfully synthesized due to the incorporation of aggregation breaking building blocks, such as pseudoprolines. Post synthesis dialysis of the purified materials resulted in correctly folded protein which was validated using circular dichroism (CD). Nonetheless, this methodology has some limitations, including low yields and the incompatibility with the introduction of C-terminal modifications which are often used in the synthesis of ABPs. 2-chlorotrityl chloride (CTC) resin on the other hand does enable the introduction of modifications to the C-terminus, however this resin is instable towards heat and results therefore in the premature resin-release of the peptide during synthesis.




**Figure 3.** Schematic representation of an ABP to covalently capture active cysteine protease enzymes followed by a pull-down.

To overcome this issue, we set out to investigate a NCL approach compatible with CTC resin for the synthesis of LC3 ABPs in **chapter 6** (Fig. 3). A two-fragments approach using a N-terminal fragment of 77 amino acids and a C-terminal fragment of 43 amino acids was explored for both LC3A and LC3B. Similar ligation strategies could be applied for both proteins, due to their high sequence similarity. Since, LC3A has one cysteine residue located in the N-terminal region of the protein and LC3B is cysteine free, a native cysteine could not be used for NCL and it was necessary to introduce an alanine to cysteine mutation in both proteins to allow a successful NCL/desulfurization strategy. Disconnection of the protein at alanine78 and glutamine77 was the preferred junction, however, deemed us to use the 3,4-diaminobenzoic acid (Dbz) linker for thioester formation.[20] Glutamine is a challenging amino acid in thioester formation due to

the possibility of self-cyclization and the Dbz linker minimizes this side reaction. The C-terminus of both proteins appeared to be challenging due to the high hydrophobicity. By using automated fast flow synthesis (AFPS) the aggregation on resin could be monitored and solved by the site specific introduction of pseudoprolines. The high hydrophobicity forced us to search for alternative purification methods, since RP-HPLC did not result in pure material. Purification via the catch-and-release method (Belyntic GmbH, Berlin, Germany) resulted in decent purity peptide material for the NCL. The N-terminus of the proteins containing a biotin enrichment tag was synthesized without difficulties. The two probes were successfully synthesized using NCL-desulfurization and used in *in vitro* pull-down assays of Atg4B. Wild-type Atg4B is successfully pulled-down by both LC3A-PA and LC3B-PA. However, the catalytic inactive Atg4B mutant (C74A) was also pulled-down by LC3B, even when harsh washing conditions were used indicating a true covalent interaction. This was surprisingly, since propargyl probes are known to selectively react with active site cysteines and are considered to be inert and unreactive towards non-active site cysteine residues.[21] A second cysteine residue was identified in the active site, located in close proximity to the active site cysteine that might complement the catalytic triad. Future work to elucidate reactivity of our probes towards ATG proteins might include protein mutagenesis to facilitate a pull-down experiment using a double mutant. Additionally, the strong interaction observed might be due to the additional length of LC3Bs C-terminus since the propargyl moiety was coupled to the full length sequence that mimics the phospholipid part found in membrane bound LC3s. To investigate this option longer or shorter LC3-PA probes, or probes containing different lipid mimics could be synthesized and tested in similar pull-down experiments. LC3-probes containing a phospholipid at the C-terminus connected with a dehydroalanine warhead would be valuable to find other conjugating enzymes for LC3 as was performed for Ub by Mulder et al.[22] In addition, the synthesis of probes containing an vinylamide (VA) could aid in the search for LC3 deconjugating enzymes.[4] The specific functions of the three LC3 isoforms is still not clear and to investigate the differences between them, LC3A-C probes containing a photo-crosslinker could be synthesized. LC3 contains a hydrophobic pocket important for the interaction with other proteins, introduction of a photo-crosslinker in this region could identify new interaction partners of the different LC3 isoforms. If different interaction partners of the three isoforms are identified this might clarify isoform function. In the past, this approach has been successfully applied on SUMO2 by Mootz and coworkers.[23] Furthermore, expansion of the LC3 toolbox could be optimized by introducing solubilizing tags to increase yield and solubility of the C-terminal fragments.

Finally, in **chapter 7** a new lysine ubiquitination method is explored with the goal to open up a route towards the easy synthesis of ubiquitinated substrates and (branched) polyUb chains. The progress made in the chemical protein synthesis field has resulted

already in the synthesis of many diUbs, polyUb chains and ubiquitinated proteins. These methods, however, often require special buildings blocks[24,25], and complex synthetic strategies[26] which are not always accessible for all biochemical labs. In addition, research intro branched Ub chains is still in its infancy due to the challenging synthesis of these complex proteins. Hence, additional methodologies would contribute to the more effective exploration of this enigmatic feature of ubiquitination. The selective lysine acylation reagents explored by Jensen et al. seemed very interesting for this purpose. These reagents are based on 2,4-dichloro-6-sulfonic acid phenol esters and at high pH the reagents tend to be regioselective for lysine acylation while at neutral pH they show high preference for N-terminal modification (Fig. 4).[27,28]

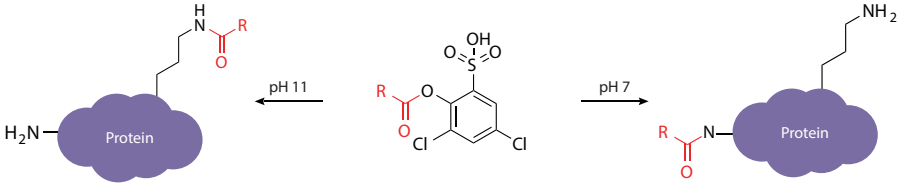


**Figure 4.** Selective modification under aqueous conditions by using 2,4-dichloro-6-sulfonic acid phenol esters.

We envisioned the described phenol esters to be suitable for the synthesis of Ub chains, ranging from dimers to polymers. To form specific Ub-linkage types only the desired lysine residue in the target should be accessible while the other lysines are protected with protective groups (such as nosyl (Ns)) during acylation. The activation of large proteins e.g. Ub with the phenol esters did not cause any difficulties and ubiquitination of model peptides proved to be selective and efficient. When we moved on to the ubiquitination of more complex peptides such as Ub itself, containing only one available lysine, a decrease in selectivity was observed. Unexpectedly acylation at alkaline pH resulted in predominant lysine acylation, but to some extend also N-terminal acylation was observed after prolonged reaction times. At neutral pH both N-terminal and lysine acylation were observed. A pilot experiment also indicated that polymerization reactions using this method gives rise to higher order polyUb chains. To develop this methodology further, future optimization studies could benefit from protecting the N-terminus to avoid over-acylation and scouting for other protective groups beside the Ns-group. If only one Ns protecting group is present on the peptide/protein this could easily be removed by using MPAA and base, however when more complex protein mixtures containing 6-7 Ns groups are used, the deprotection is more challenging. Complete deprotection was only observed when using harsh conditions including heating to 60 °C for three days. Exchange of the Ns protecting group for another protecting group, such as azidolysine might be an interesting alternative. Reduction of azidolysine to lysine is easily performed using the protein friendly reducing reagent TCEP. Ub derivatives containing two or more orthogonal protecting groups could be synthesized to investigate

the possibilities to synthesize branched chains. This branched Ub chain methods could possibly be expanded into the synthesis of the underexplored hybrid chains, where both Ub and Ubl-proteins are conjugated into one chain.[29] Furthermore, this methodology could give easy access to ubiquitinated peptides and proteins for biological labs, which with the current approaches is not feasible. To conclude, the work described in this chapter proves promising for expansion into the controlled synthesis of (branched) polyUb chains, which could eventually be a valuable asset in the deciphering of cellular signaling by the still underexplored branched Ub chains.

Collectively, the work described in this PhD thesis focusses on the chemical protein synthesis of a diverse set of chemical tools and assay-reagents to facilitate future studies of Ub(l) biology. Although a wide variety of Ub-chains and tools have been prepared chemically, translating such protocols to other Ubls requires careful fine-tuning of conditions and used approaches. Non-the-less tailored SPPS and ligation techniques have enabled the synthesis of novel probes based on GFP-nanobody, SUMO and LC3s. The methods and produced tools described in this thesis might form important additions to the Ub(l) chemical toolbox and are expected to contribute in the further deciphering of the Ub(l) code.

## References


1. Y. Tan, H. Wu, T. Wei, X. Li, *J. Am. Chem. Soc*. **2020**, 20288–20298.
2. A. N. Hartrampf, A. Saebi, M. Poskus, Z. P. Gates, A. J. Callahan, A. E. Cowfer, S. Hanna, S. Antilla, C. K. Schissel, A. J. Quartararo, X. Ye, A. J. Mijalis, M. D. Simon, A. Loas, S. Liu, C. Jessen, T. E. Nielsen, B. L. Pentelute, *Science*. **2020**, 987, 980–987.
3. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science*. **1994**, 266, 776–779.
4. M. P. C. Mulder, R. Merkx, K. F. Witting, D. S. Hameed, D. El Atmioui, L. Lelieveld, F. Liebelt, J. Neefjes, I. Berlin, A. C. O. Vertegaal, H. Ovaa, *Angew. Chemie - Int. Ed*. **2018**, 57, 8958–8962.
5. B. T. Xin, J. Gan, D. J. Fernandez, K. P. Knobeloch, P. P. Geurink, H. Ovaa, *Org. Biomol. Chem*. **2019**, 17, 10148–10152.
6. K. F. Witting, G. J. van der Heden van Noort, C. Kofoed, C. Talavera Ormeño, D. el Atmioui, M. P. C. Mulder, H. Ovaa, *Angew. Chemie - Int. Ed*. **2018**, 57, 14164–14168.
7. J. Wesolowski, V. Alzogaray, J. Reyelt, M. Unger, K. Juarez, M. Urrutia, A. Cauerhff, W. Danquah, B. Rissiek, F. Scheuplein, N. Schwarz, S. Adriouch, O. Boyer, M. Seman, A. Licea, D. V. Serreze, F. A. Goldbaum, F. Haag, F. Koch-Nolte, *Med. Microbiol. Immunol.* **2009**, 198, 157–174.
8. S. Muyldermans, *Annu. Rev. Biochem*. **2013**, 82, 775–797.
9. E. Y. Yang, K. Shah, *Front. Oncol*. **2020**, 10.
10. H. D. Herce, D. Schumacher, A. F. L. Schneider, A. K. Ludwig, F. A. Mann, M. Fillies, M. A. Kasper, S. Reinke, E. Krause, H. Leonhardt, M. C. Cardoso, C. P. R. Hackenberger, *Nat. Chem*. **2017**, 9, 762–771.
11. R. W. Cheloha, T. J. Harmand, C. Wijne, T. U. Schwartz, H. L. Ploegh, *J. Biol. Chem*. **2020**, 295, 15307–15327.
12. V. De Filippis, F. De Antoni, M. Frigo, P. P. De Laureto, A. Fontana, *Biochemistry.* **1998**, 37, 1686–1696.
13. A. Flotho, F. Melchior, *Annu. Rev. Biochem*. **2013**, 82, 357–385.
14. D. Shin, R. Mukherjee, D. Grewe, D. Bojkova, K. Baek, A. Bhattacharya, L. Schulz, M. Widera, A. R. Mehdipour, G. Tascher, P. P. Geurink, A. Wilhelm, G. J. van der Heden van Noort, H. Ovaa, S. Müller, K. P. Knobeloch, K. Rajalingam, B. A. Schulman, J. Cinatl, G. Hummer, S. Ciesek, I. Dikic, *Nature.* **2020**, 587, 657–662.
15. T. E. T. Mevissen, M. K. Hospenthal, P. P. Geurink, P. R. Elliott, M. Akutsu, N. Arnaudo, R. Ekkebus, Y. Kulathu, T. Wauer, F. El Oualid, S. M. V. Freund, H. Ovaa, D. Komander, *Cell* **2013**, 154, 169.
16. N. Hattersley, L. Shen, E. G. Jaffray, R. T. Hay, *Mol. Biol. Cell.* **2011**, 22, 78–90.
17. N. Wesch, V. Kirkin, V. V. Rogov, *Cells.* **2020**, 9, 1–25.
18. Y. K. Lee, J. A. Lee, *BMB Rep*. **2016**, 49, 424–430.
19. N. Mizushima, B. Levine, A. M. Cuervo, D. J. Klionsky, *Nature.* **2008**, 451, 1069–1075.
20. J. B. Blanco-Canosa, P. E. Dawson, *Angew. Chemie - Int. Ed*. **2008**, 47, 6851–6855.
21. R. Ekkebus, S. I. Van Kasteren, Y. Kulathu, A. Scholten, I. Berlin, P. P. Geurink, A. De Jong, S. Goerdayal, J. Neefjes, A. J. R. Heck, D. Komander, H. Ovaa, *J. Am. Chem. Soc*. **2013**, 135, 2867–2870.
22. M. P. C. Mulder, K. Witting, I. Berlin, J. N. Pruneda, K. P. Wu, J. G. Chang, R. Merkx, J. Bialas, M.

Groettrup, A. C. O. Vertegaal, B. A. Schulman, D. Komander, J. Neefjes, F. El Oualid, H. Ovaa, *Nat. Chem. Biol*. **2016**, 12, 523–530.

23. K. Brüninghoff, A. Aust, K. F. Taupitz, S. Wulff, W. Dörner, H. D. Mootz, *ACS Chem. Biol*. **2020**, 15, 2406–2414.
24. B. T. Xin, B. D. M. Van Tol, H. Ovaa, P. P. Geurink, *Org. Biomol. Chem*. **2018**, 16, 6306–6315.
25. G. J. Van Der Heden Van Noort, R. Kooij, P. R. Elliott, D. Komander, H. Ovaa, *Org. Lett*. **2017**, 19, 6490–6493.
26. H. Sun, A. Brik, *Acc. Chem. Res*. **2019**, 52, 3361–3371.
27. K. B. Jensen, J. H. Mikkelsen, S. P. Jensen, S. Kidal, G. Friberg, T. Skrydstrup, M. B. F. Gustafsson, *Bioconjug. Chem*. **2022**, 33, 172–179.
28. J. H. Mikkelsen, M. B. F. Gustafsson, T. Skrydstrup, K. B. Jensen, *Bioconjug. Chem*. **2022**, 33, 625–633.
29. D. A. Pérez Berrocal, K. F. Witting, H. Ovaa, M. P. C. Mulder, *Front. Chem*. **2020**, 7, 1–9.